



Biodiversity Image Quality Metadata Augments Convolutional Neural Network Classification of Fish Species

Jeremy Leipzig¹ , Yasin Bakis² , Xiaojun Wang² , Mohannad Elhamod³ ,
Kelly Diamond^{4,5} , Wasila Dahdul⁶ , Anuj Karpatne³ , Murat Maga^{4,5} ,
Paula Mabee⁷ , Henry L. Bart Jr.² , and Jane Greenberg¹ 

¹ College of Computing and Informatics, Metadata Research Center, Drexel University, Philadelphia, PA 19104, USA

jnl47@drexel.edu

² Tulane University Biodiversity Research Institute, 3705 Main St., Building A3, Belle Chasse, LA 70037, USA

³ Department of Computer Science, Virginia Tech, Rice Hall Information Technology Engineering Building, 22904, 85 Engineer's Way, Charlottesville, VA 22903, USA

⁴ Division of Craniofacial Medicine, Department of Pediatrics, University of Washington, Seattle, WA, USA

⁵ Center for Development Biology and Regenerative Medicine, Seattle Children's Hospital, Seattle, WA, USA

⁶ Department of Digital Services, UC Irvine Libraries, Irvine, USA

⁷ National Ecological Observatory Network Program, Battelle, Boulder, CO, USA

Abstract. Biodiversity image repositories are crucial sources for training machine learning approaches to support biological research. Metadata about object (e.g. image) quality is a putatively important prerequisite to selecting samples for these experiments. This paper reports on a study demonstrating the importance of *image quality metadata* for a species classification experiment involving a corpus of 1935 fish specimen images which were annotated with 22 metadata quality properties. A small subset of high quality images produced an F1 accuracy of 0.41 compared to 0.35 for a taxonomically matched subset low quality images when used by a convolutional neural network approach to species identification. Using the full corpus of images revealed that image quality differed between correctly classified and misclassified images. We found anatomical feature visibility was the most important quality feature for classification accuracy. We suggest biodiversity image repositories consider adopting a minimal set of image quality metadata to support machine learning.

Keywords: Image classification · Convolutional neural networks · Image metadata · Quality metadata

1 Introduction

1.1 Quality Metadata for Species Image Repositories

The extensive growth in open science repositories, and, in particular, the underlying application of rich metadata has potential value for data mining, machine learning and deep learning (ML/DL). Research confirms value of metadata for machine learning, automatic document classification [1], and reproducible research pipelines [2, 3]. Less common, but of paramount importance is metadata that denotes the quality of the object being represented. Metadata addressing quality control characteristics of data can support the data cleaning steps common to virtually all ML/DL analyses. Computer vision offers proof with quality-specific metadata that's important for selecting, training, and validation and test image sets. For example, Ellen et. al found the use of context metadata, consisting of hydrographic, geotemporal, and geometric data, representing plankton images improved the accuracy of a convolutional neural network (CNN) classifier [4]. Tang found a 7% gain in mean average precision after including GPS coordinates in a general image classification task [5]. These studies shed light on an important area of metadata research that has broad implications for leveraging collections of digital images across nearly every scientific discipline.

Of particular interest is biological specimen image collections, given their value as a data source for species identification and morphological study [6, 7]. The research presented in this paper addresses this topic in the context of an NSF supported Harnessing the Data Revolution (HDR) project, Biology-Guided Neural Networks for Discovering Phenotypic Traits (BGNN). A team of information and computer scientists, biologists, and image experts are collaborating to develop a novel set of artificial neural networks (ANNs) for classifying fish species and extracting data on fish external morphological features from images of fish specimens. Unlike genomic data, specimen trait data is largely unstructured and not machine readable. The paucity of usable trait data for many groups of organisms has led to efforts to apply neural network-based classification and morphological analysis to the extensive store of existing species photographic images to automatically extract trait data in an unsupervised manner. The product of these efforts, the focus of BGNN, should improve our ability to derive phenotype data from digital analogs of specimens. Image quality metadata is recognized as an important factor in the selection of images for ML/DL.

1.2 Image Metadata Content Description and Quality

Over the last few decades a number of metadata schemes have been developed that describe specimens found in large scale digital specimen repositories. These metadata standards focus chiefly on description, with limited attention to image quality. The MODAL framework [8, 9] provides a mechanism for understanding the range and diversity of these metadata standards with their domain foci (general to specific) and limited number of image quality properties. Metadata for analog images, our immediate focus, falls into two classes: *descriptive metadata*, generated by a curator; and *technical metadata*, automatically generated by technology capturing the image.

The Dublin Core, Darwin Core, and the Audubon Core offer properties that accommodate quality at a high level. The more technically oriented metadata standards identified in Table 1 provide a richer set of image quality properties, although they need to be measured against parameters defining quality.

Table 1. Example technical and biomedical metadata standards

Metadata standard	Primary focus	Metadata quality property
PREMIS	Long term preservation	Fixity
EXIF	Image formats	X/Y dimensions, compression, color space
DICOM [10]	Medical imaging	imageQuality (1–100)

Semantic ontologies and controlled vocabularies can also be used to indicate value. Table 2 identifies two ontologies, and example semantics describing image quality.

Table 2. Selected ontologies

Ontology	Primary focus	Semantics/metadata values
Biomedical Image Ontology (BIM) [11]	Biomedical images	Image filters, ImagePreProcessing, ImagePostProcessing
OntoNeuroBase [12]	Neuro imaging	Structure of interest, orientation, segmentation result

Overall, the schemes identified here vary in their coverage of content description for discovery, provenance tracking, and technical aspects that can help determine aspects of quality. Our assessment finds there does not yet exist a targeted metadata standard that captures the specimen image quality. This limitation motivates our research purpose and goals.

1.3 Purpose and Goals

The overall aim of our research was to examine the importance of image quality metadata for species classification. Our goals were the following: 1) Determine if the annotated image quality affected classification accuracy. 2) Determine which specific quality annotations were most important for classification accuracy. 3) Make recommendations for future image quality metadata in large image repositories.

2 Methods

To address the above goals, we conducted an empirical analysis as described below.

2.1 Sample

A sample of 23,807 digital images of fish specimens was obtained from the Illinois Natural History Survey (INHS) Fish Collection, from the Great Lakes Invasives Network (GLIN) Project [13]. Duplicate images were removed, and file formats, institution code, catalog numbers and suffixes to file names, the images were transferred to a BGNN file server for analysis. Specimen collection information (occurrence records) for the images were gathered from FishNet2 [14] and the scientific names were updated using Eschmeyer's Catalog of Fishes [15].

Next, we established a set of metadata properties to record image quality for the digitized fish specimens (Suppl. Table 1). The set of properties is based on the expertise of informaticians, fish experts, and data entry technicians at Tulane University's Biodiversity Research Institute, with feedback from members of the Metadata Research Center, Drexel University. The scheme includes 22 metadata properties, requiring the content-value of a categorical concept, free text, a Boolean operator, or a score. A web-based form, and an underlying SQL-based database help to expedite capturing the metadata content (Fig. 1).

2.2 Descriptive Statistical Analysis of Quality

A basic exploratory data analysis was performed on quality metrics. Quality averages by taxonomic groups (genus and species) were examined in order to understand potential biases.

2.3 Implementation of a CNN-Based Classification Pipeline

A convolutional neural network image classification pipeline was developed using PyTorch [16] with Torchvision [17] extensions. Genera (genus groups) and species (genus + specific epithet combinations) were trained and inferred simultaneously using a novel multi-classifier model, called a Hierarchy-Guided Neural Network (in submission). Several hyperparameters, including learning rate, regularization lambda, early stopping patience were tuned prior to this quality analysis.

2.4 Classification Accuracy Using High vs Low Quality Subsets

Using the composite median image_quality score of 8, we divided the data set into low-quality and high-quality subsets. Some species are inherently more visually similar to others, so in a classification scenario, an unbalanced distribution of taxa would confound

2.5 Quality Features Distinguishing Correctly and Incorrectly Identified Species

Image Metadata Entry



ILLINOIS NATURAL HISTORY SURVEY
 1. *Perca flavescens*
 2. Yellow perch
 3. Grand River, Illinois, USA
 4. 10 cm
 5. 1999
 6. 1999
 7. 1999
 8. 1999
 9. 1999
 10. 1999

8. Does the image have a color scale bar?

9. Is there a specimen label for each fish specimen in the image?

10. Does the image contain any objects other than fish, ruler, specimen label, color bar?

11. Do any of the objects in the image overlap or hide any parts of the fish?

12. One a clock face (click "show-hide clock"), which direction is the fish's head pointing?

Size:

Image Name:

Image Size: width: px

Image Size: height: px

13. Which surfaces of the specimen are visible (left side visible when head points left)?

14. Is the body of the fish specimen straight or curved?

15. Are any parts of the fish cropped from view?

16. Are all parts (including fins) of the fish visible?

17. Are any of the fins lying down or folded oddly?

18. Rate the brightness of the fish specimen.

19. Is the background of the image uniform?

20. Is the fish specimen in focus?

21. Does the specimen have any color issues?

22. Rate overall image quality from 1-10:

1. Is a fish specimen visible in the image?

2. How many fish specimens are in the image?

3. Does the image include a ruler?

4. Does the image include a specimen label?

5. Does the number in the image name INHS_FISH_9999.jpg match the number on the label?

6. Is the catalog number on the label numerical or some other format (e.g., a barcode)?

7. Is the specimen label detailed (multiple lines of information) or does it have just a specimen number?

1. ☐ Yes ☐ No

2. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

3. ☐ Yes ☐ No

4. ☐ Yes ☐ No

5. ☐ Yes ☐ No

6. ☐ Yes ☐ No

7. ☐ Detailed ☐ Basic

At the time of this publication, metadata annotations indicating quality have been created for a total of 1935 images. In order to prepare images for classification, Advanced Normalization Tools in R (ANTsR) [18] was used to subtract background shading, rulers, and labels using a supervised segmentation with a training set developed in 3D Slicer [19]. Our analysis focused on comparing low and high-quality images that were roughly balanced by genus and species composition, in order to control for the effect of inherent differences in identification difficulty that vary among taxa. We noted that image quality varied non-randomly between species, perhaps due to batch effects as well as anatomic differences between fish taxa that affect photographic fidelity (Fig. 2).



Fig. 2. Examples of very low (1) and very high (10) quality images of *Esox americanus*

3 Results

3.1 Low/High Subset Comparison

A t-test of F1 scores generated by several runs on the small balanced high and low quality subsets showed a small but significant difference in accuracy (0.41 vs 0.35, $p\text{-val} = 0.031$) (Fig. 3).

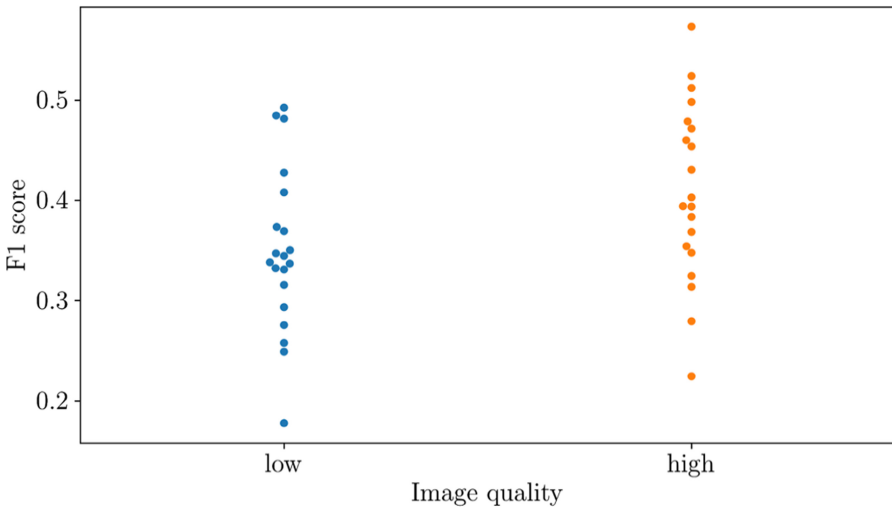


Fig. 3. F1 test score across 19 trials on genus classification using low quality (mean 0.35) and high quality (mean 0.41). Using high quality images produced better F₁ scores (0.41 vs 0.35, $p\text{-val} = 0.031$).

3.2 Quality by Classification Outcome

Here we compared correctly classified vs misclassified images using a test set of 341 images (278 correctly classified and 63 misclassified). A confusion matrix (Suppl. Fig. 3) shows that most misclassifications occur between species within the same genus, although these misclassifications are not symmetric.

Comparing the means of quality scores between correctly classified images reveals five quality features correlated with classification accuracy: *if_curved*, *if_parts_visible*, *if_overlapping*, and *image_quality*, and two negatively correlated: *if_background_uniform*, and *if_fins_folded*. While *image_quality* is the strongest variable, a logistic regression which includes all features except *image_quality* (to avoid collinearity), reveals *if_parts_visible* (p-val = 0.0001) as the sole significant covariate. (Suppl. Table 2).

4 Discussion

In this paper we show that image quality measures impact classification accuracy. This was demonstrated using two approaches - a dichotomous split of the image corpus using the manually-annotated *image_quality* metric (Suppl. Fig. 2), and a comparison of correctly and incorrectly classified images from the entire quality-annotated data set. Our abilities to discern the importance of quality are hampered by three factors: 1) a relative paucity of low-quality images in our dataset, 2) the nature of classification - some fish are simply more similar to their brethren (Suppl. Fig. 3) - but we have attempted to control for this where possible, and 3) some taxa are inherently more difficult to position or illuminate for photography. Although image quality was high overall within this collection, it varied substantially, and perhaps more importantly, unevenly with respect to taxa within image repositories, which may belie both individual variation in photography and batch effects associated with submitters (Fig. 4). Our results lead to a number of recommendations for assessing quality of images from biodiversity image repositories to support machine learning analyses.

The *quality_score* assigned by curators, while based on a rubric, does lend itself to some inter-rater error. We surmised that a composite metric of the binary quality items (e.g. *if_curved*, *if_fins_folded*, etc..) could represent a more objective score, and explored this, but it ultimately did not prove substantially better than "image_quality".

The quality scores generated by our curators included some that are strictly technical (blur, color issues), those that would apply to any biodiversity catalog (*if_parts_visible*) and those that are specific to fish (e.g. *if_fins_folded*). We contend that all three types of quality (technical, biospecimen, taxon-specific) are important to include for biorepositories. The automated measurement of technical image quality, and possible higher-level judgments, can help accelerate the collection of this metadata. Local features [20, 21] such as fin-level textures that would indicate lepidotrichia and global features such as large segmented areas and basic image characteristics of color and shape are logically distinct from semantic quality judgments made by the curators in this project ("folded fins"/"label obstruction"), though automated semantic quality annotations are within the capabilities of neural networks.

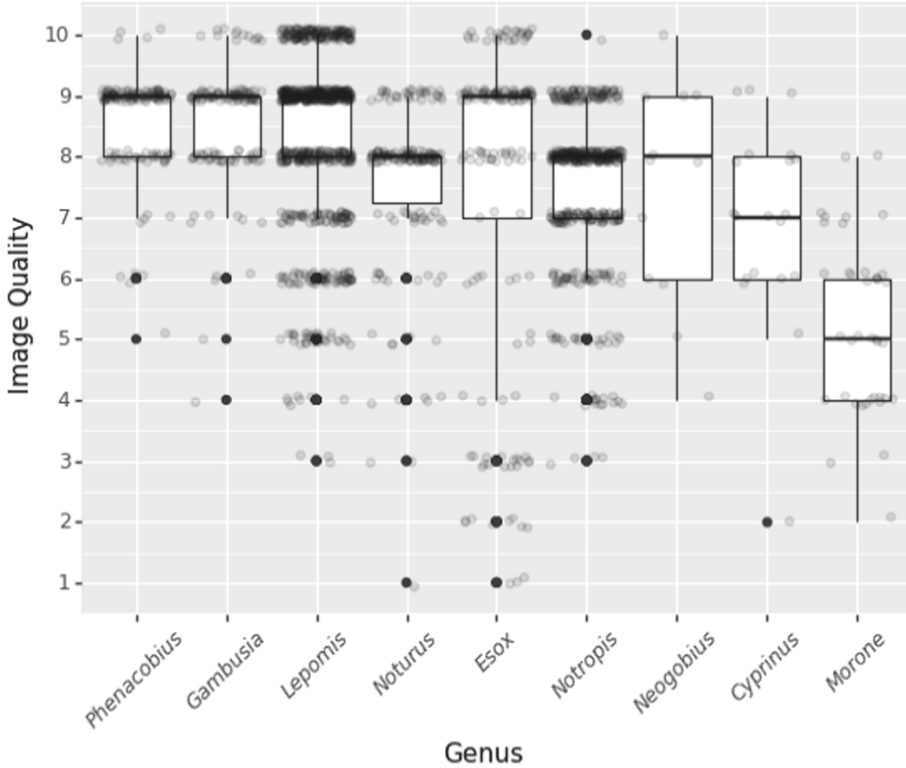


Fig. 4. Quality scores by genera with more than 10 individuals clearly show variability within and between genera.

We suggest repositories enforce provenance tracking metadata to assist in identifying batch effects or other confounds introduced by disparate labs photographing specimens with different settings and equipment.

We observed that the classification task is hampered by its dependence on accuracy instead of more direct intermediate measures, for example, the number of features detected. Certain types of low-quality images may serve to augment robustness to classify real-world specimens - a technique called “noise injection”. These uses suggest annotating for quality, rather than simply culling, is a preferable strategy. Quality metadata to aid robustness and generalizability in machine learning, rather than a narrow focus on pristine specimens, is an open area for future work.

5 Conclusion

The main objective of this research was to determine if annotated image quality metadata impacted generic and species-level classification accuracy of a convolutional neural network. We conducted an empirical analysis to examine which specific quality annotations, based on 22 metadata properties, were important for classification accuracy.

Our key finding was that images with high-quality metadata metrics had significantly higher classification accuracy in convolutional neural network analysis than images with low-quality metadata metrics.

We offer a number of recommendations for assessing the quality of images from biodiversity image repositories to support machine learning analyses. This investigation serves as a baseline study of useful metadata for assessing image quality. The methodology, our approach, and the base-level scheme of 22 metadata properties, serves to inform other research that seek to record image quality; and, further study the impact on classification for other fishes, other specimens, and even other disciplines where the image is a central object. Overall the research conducted serves the needs of the BGNN project, and biologically-focused, machine-learning projects generally, for determining whether images for biodiversity specimen image repositories are useful for higher-level analysis.

6 Supplemental Materials

Supplementary materials are available at <https://doi.org/10.6084/m9.figshare.13096199>.

Raw data is available at <https://bgnn.org/INHS>. Reproducible code is available at <https://github.com/hdr-bgnn/iqm>.

Acknowledgements. Research supported by NSF OAC #1940233 and #1940322.

We acknowledge Tulane University Technology Services for server set-up and access supporting this research; Justin Mann, Aaron Kern, Jake Blancher for curation work; and Chris Taylor, Curator of Fishes and Crustaceans, INHS, for providing the fish specimen images and metadata (INHS support: NSF Award #1400769).

Competing Interests. The authors have declared that no competing interests exist.

References

1. Han, H., Giles, C.L., Manavoglu, E., et al.: Automatic document metadata extraction using support vector machines. In: 2003 Joint Conference on Digital Libraries, 2003. Proceedings, pp. 37–48 (2003)
2. Schelter, S., Boese, J.-H., Kirschnick, J., et al.: Automatically tracking metadata and provenance of machine learning experiments. In: Machine Learning Systems workshop at NIPS (2017)
3. Leipzig, J., Nüst, D., Hoyt, C.T., et al.: The role of metadata in reproducible computational research. arXiv [cs.DL] (2020)
4. Ellen, J.S., Graff, C.A., Ohman, M.D.: Improving plankton image classification using context metadata. *Limnol. Oceanogr. Methods* **2**, 17 (2019)
5. Tang, K., Paluri, M., Fei-Fei, L., et al.: Improving image classification with location context. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1008–1016 (2015)
6. Borges, L.M., Reis, V.C., Izbicki, R.: Schrödinger’s phenotypes: herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Methods Ecol. Evol.* (2020). <https://doi.org/10.1111/2041-210X.13450>

7. Hernández-Serna, A., Jiménez-Segura, L.F.: Automatic identification of species with neural networks. *PeerJ* **2**, e563 (2014)
8. Willis, C., Greenberg, J., White, H.: Analysis and synthesis of metadata goals for scientific data. *J Am Soc Inf Sci Technol* **63**, 1505–1520 (2012)
9. Greenberg, J.: Understanding metadata and metadata schemes. *Cataloging Classif. Q.* **40**, 17–36 (2005)
10. Bidgood, W.D., Jr., Horii, S.C.: Introduction to the ACR-NEMA DICOM standard. *Radio-graphics* **12**, 345–355 (1992)
11. Bukhari, A.C., Nagy, M.L., Krauthammer, M., et al.: BIM: an open ontology for the annotation of biomedical images. In: ICBO (2015)
12. Temal, L., Lando, P., Gibaud, B., et al.: OntoNeuroBase: a multi-layered application ontology in neuroimaging. In: Second Workshop: Formal Ontologies Meet Industry (FOMI 2006) (2006)
13. Great Lakes Invasives Network Home. <https://greatlakesinvasives.org/portal/index.php>. Accessed 2 Aug 2020
14. FishNet2. <https://www.fishnet2.net/>. Accessed 2 Aug 2020
15. Eschmeyer's Catalog of Fishes. <https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes>. Accessed 2 Aug 2020
16. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., et al. (eds.) *Advances in Neural Information Processing Systems*, vol. 32, pp 8026–8037. Curran Associates, Inc. (2019)
17. Marcel, S., Rodriguez, Y.: Torchvision the machine-vision package of torch. In: In: Proceedings of the 18th ACM international conference on Multimedia, pp. 1485–1488. Association for Computing Machinery, New York (2010)
18. Avants, B.B., Kandel, B.M., Duda, J.T., et al.: ANTsR: ANTs in R: quantification tools for biomedical images. R package version 0542 (2019)
19. Kikinis, R., Pieper, S.D., Vosburgh, K.G.: 3D slicer: a platform for subject-specific image analysis, visualization, and clinical support. In: Jolesz, F.A. (ed.) *Intraoperative Imaging and Image-Guided Therapy*, pp. 277–289. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-7657-3_19
20. Shyu, C.R., Brodley, C.E., Kak, A.C., et al.: Local versus global features for content-based image retrieval. In: *Proceedings, IEEE Workshop on Content-Based Access of Image and Video Libraries* (Cat. No.98EX173), pp 30–34 (1998)
21. Vogel, J., Schwaninger, A., Wallraven, C.: Categorization of natural scenes: local vs. global information. In: *Proceedings of Conference Army Physicians Cent Mediterr Forces* (2006)